On the parametric maximum likelihood estimator for independent but non-identically distributed observations with application to truncated data*

F. Leroy[1,2,*], J.Y. Dauxois[3], P. Tubert-Bitter[1,2]

[1] *Inserm, CESP Centre for research in Epidemiology and Population Health, U1018, Biostatistics Team, F-94807, Villejuif, France;* [2] *Univ Paris-Sud, UMRS1018, F-94807, Villejuif, France;* [3] *Université de Toulouse-INSA, IMT UMR CNRS 5219, Toulouse, France*

Abstract

We investigate the parametric maximum likelihood estimator for truncated data when the truncation value is different according to the observed individual or item. We extend Lehmann's proof (1983) of the asymptotic properties of the parametric maximum likelihood estimator in the case of independent non-identically distributed observations. Two cases are considered: either the number of distinct probability distribution functions that can be observed in the population from which the sample comes from is finite or this number is infinite. Sufficient conditions for consistency and asymptotic normality are provided for both cases.

Keywords: Parametric maximum likelihood estimator; Independent non-identically distributed observations; Consistency; Asymptotic normality; Truncated data.

1 Introduction

Truncated data arise frequently in survival analysis or in astronomy. For instance, left-truncated data occur when one wants to estimate the luminosity of an astronomical object but one can only detect astronomical objects which are sufficiently bright (Woodroofe, 1985).

*This work was supported by the Fondation ARC (fellowship DOC20121206119 to Fanny Leroy).

*Corresponding author. Email address: fanny.leroy@inserm.fr (F. Leroy)

In survival studies, data on time to onset of a disease that are collected retrospectively in case registries are right-truncated because the sample includes only the cases, with their information, that have already occurred and not the cases that may occur in the future. A well-known example arised in the late eighties with the estimation issue of AIDS incubation time distribution for transfusion-induced AIDS (Lagakos et al., 1988). In the same way, data on incubation period of inhalational anthrax, from people living or working in a city of Russia where an anthrax outbreak occurred in April 1979 and where a subsequent public health intervention was led, are right-truncated because the public health intervention prevented some deaths (Brookmeyer et al., 2001). In the last both cases, accurate estimation of the distribution of the incubation time could help in setting up public health policies to diagnose earlier or improve the treatment of infected people.

Let X be the random variable of interest, the luminosity of an astronomical object or the incubation time of a disease in our examples. We assume a parametric model for the distribution of the random variable X and the maximum likelihood estimator is considered. Let θ be the vector of unknown parameters of the assumed parametric model. Let $F(.;\theta)$ be the cumulative distribution function of X and $f(.;\theta)$ its probability distribution function. Let t belonging to $\mathbb{R}$. The random variable X is said to be right-truncated (resp. left-truncated) by the truncation value t when the variable X is observed only if its realization is smaller (resp. larger) than t. The truncation value t may be different according to the individual or item. Let (x_1,t_1), (x_2,t_2), $\ldots$, (x_n,t_n) be the n truncated observations, where x_i is the realization of X and t_i is the truncation value. All observed data meet the condition $x_i \leq t_i$ in the case of right-truncated data or the condition $x_i \geq t_i$ in the case of left-truncated data. Right-truncated (resp. left-truncated) data on X consist of independent realizations of random variables with respective distribution the conditional distribution of X_i given $\{X_i \leq t_i\}$ (resp. $\{X_i \geq t_i\}$), that is with cumulative distribution function $F(.;\theta)/F(t_i;\theta)$ (resp. $F(.;\theta)/\left(1-F(t_i;\theta)\right)$) and probability distribution function $f_i(.;\theta) = f(.;\theta)/F(t_i;\theta)$ (resp. $f_i(.;\theta) = f(.;\theta)/\left(1-F(t_i;\theta)\right)$). Consequently, if the truncation value is different according to the individual or item, truncated data consist of independent but non-identically distributed observations. In this paper, we deal with this case.

Asymptotic properties of the parametric maximum likelihood estimator for independent and identically distributed observations in the multiparameter case have been explored by Chanda (1954) and Lehmann (1983). Chanda (1954) solved the normal equations to prove the consistency of this estimator whereas Lehmann (1983) studied the sign of a function of the log-likelihood on a sphere with center the true value of the parameter vector. While Bradley and Gart (1962) developed the extension of the proof of Chanda (1954) for independent but non-identically distributed observations, there is no extension of the proof of Lehmann (1983). There are some other proofs of asymptotic properties like the proof using empirical processes and exposed by Van der Vaart and Wellner (2000) that yields to different statements of assumptions that may not be easy to verify in specific situations. In the present article, we develop the extension of the proof of Lehmann (1983) in the case of independent but non-identically distributed observations. In their paper, Bradley and Gart (1962) considered two cases: either the number of distinct probability distribution functions that can be observed in the population from which the sample comes from is finite or this number is infinite. For the sake of generality, we consider these both cases. In the case of an infinite number of distinct probability distribution functions, the assumptions that are sufficient conditions for consistency and asymptotic normality of the parametric maximum likelihood estimator are slighlty different than in the paper of Bradley and Gart (1962). In the remaining of this paper are presented the assumptions, the theorems and the proofs of the asymptotic properties of the parametric maximum likelihood estimator.

2 Asymptotic properties

In this paper, for a sequence of random variables with index n, the convergence in probability is written $\underset{n\to+\infty}{\overset{P}{\longrightarrow}}$ and the convergence in distribution is written $\underset{n\to+\infty}{\overset{d}{\longrightarrow}}$.

2.1 Infinite number of distinct probability distribution functions

Let $(x_1, x_2, \ldots, x_n)$ be the observations of n independent random variables with respective and not necessarily identical probability distribution functions $f_i(.;\theta)$, for $i = 1, \ldots, n$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_r)$ is a vector of unknown parameters shared by all these random variables. The vector θ belongs to Θ, an open subset of $\mathbb{R}^r$. Let $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_j^0, \ldots, \theta_r^0)$ be the true value of the parameter. Let $S_i \subset \mathbb{R}$ be the support of the probability distribution function $f_i(.;\theta)$. The support S_i must be independent of the vector of unknown parameters θ. As well-known, the likelihood of the sample is written $L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^n f_i(x_i; \theta)$ and the maximum likelihood estimator is defined as $\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmax}} \; L(x_1, x_2, \ldots, x_n; \theta)$. The normal equations are

$$\nabla_\theta \log \, L(x_1, x_2, \ldots, x_n; \theta) = 0,$$

where ∇_θ is the gradient operator.

Remark 1. We assumed that the unknown parameter vector is shared by all the densities because it is the case for truncated data. However, theorems and proofs remain valid when it is not the case.

Let us introduce a set of sufficient conditions for the following theorems.

Assumption 1. The maximum likelihood estimator is solution of the normal equations.

Assumption 2. The normal equations have an unique root.

Assumption 3. For all $\theta \in \Theta$, $i = 1, \ldots, n$ and $(j, p, q) \in \{1, \ldots, r\}^3$, the partial derivatives

$$\frac{\partial \log \, f_i(.;\theta)}{\partial \theta_j}, \; \frac{\partial^2 \log \, f_i(.;\theta)}{\partial \theta_j \partial \theta_p} \; \text{ and } \; \frac{\partial^3 \log \, f_i(.;\theta)}{\partial \theta_j \partial \theta_p \partial \theta_q}$$

exist for almost all x.

Assumption 4. For all $\theta \in \Theta$, $i = 1, \ldots, n$ and $j \in \{1, \ldots, r\}$, the partial derivative

$\frac{\partial}{\partial \theta_j} f_i(.;\theta)$ is an integrable function on S_i and

$$\int_{S_i} \frac{\partial}{\partial \theta_j} f_i(x;\theta)dx \;=\; \frac{\partial}{\partial \theta_j} \int_{S_i} f_i(x;\theta)dx.$$

Assumption 5. For all $\theta \in \Theta$, $i = 1, \ldots, n$ and $(j,p) \in \{1, \ldots, r\}^2$, the partial derivative $\frac{\partial^2}{\partial \theta_j \partial \theta_p} f_i(.;\theta)$ is an integrable function on S_i and

$$\int_{S_i} \frac{\partial^2}{\partial \theta_j \partial \theta_p} f_i(x;\theta)dx \;=\; \frac{\partial}{\partial \theta_j} \int_{S_i} \frac{\partial}{\partial \theta_p} f_i(x;\theta)dx.$$

Assumption 6. For all $\theta \in \Theta$ and $j \in \{1, \ldots, r\}$,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j} \xrightarrow[n\to+\infty]{P} \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j}\right) = 0.$$

Remark 2. Assumption 4 implies that the expectation $\mathrm{E}\left(\partial \mathrm{log} f_i(X_i;\theta)/\partial \theta_j\right)$ exists and is naught for all $i = 1, \ldots, n$ and $j \in \{1, \ldots, r\}$. Consequently, the limit on the right-hand part in Assumption 6 exists.

Assumption 7. For all $\theta \in \Theta$, $i = 1, \ldots, n$ and $(j,p) \in \{1, \ldots, r\}^2$,

$$\mathrm{E}\left(\frac{\partial^2 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p}\right) \text{ and } \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^2 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p}\right)$$

exist and

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p} \xrightarrow[n\to+\infty]{P} \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^2 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p}\right).$$

Assumption 8. For all $\theta \in \Theta$, $i = 1, \ldots, n$ and $(j,p,q) \in \{1, \ldots, r\}^3$,

$$\mathrm{E}\left(\frac{\partial^3 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p \partial \theta_q}\right) \text{ and } \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^3 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p \partial \theta_q}\right)$$

exist and

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p \partial \theta_q} \xrightarrow[n\to+\infty]{P} \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^3 \mathrm{log} f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p \partial \theta_q}\right).$$

Remark 3. The weak law of large numbers (Feller, 1968) give sufficient conditions for the convergences in probability in Assumptions 6-8.

Assumption 9. There exists M such that for all $\theta \in \Theta$ and $(j, p, q) \in \{1, \ldots, r\}^3$,

$$\left| \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^3 \log f_i(X_i; \theta)}{\partial\theta_j \partial\theta_p \partial\theta_q} \right) \right| < M.$$

Assumption 10. The matrix $I\left(\theta^0\right) = \left(I_{jp}\left(\theta^0\right)\right)_{1 \leq j,p \leq r}$, where

$$I_{jp}\left(\theta^0\right) = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} E\left(- \left. \frac{\partial^2 \log f_i(X_i; \theta)}{\partial\theta_j \partial\theta_p} \right|_{\theta^0} \right),$$

is positive definite.

Remark 4. From Assumption 5 and Assumption 7, we were already sure that $I\left(\theta^0\right)$ exists and is a positive semi-definite matrix.

Assumption 11. For all $\epsilon > 0$,

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[\sum_{j=1}^{r} \left(\left. \frac{\partial \log f_i(X_i; \theta)}{\partial\theta_j} \right|_{\theta^0} \right)^2 I\left\{ \left(\sum_{j=1}^{r} \left(\left. \frac{\partial \log f_i(X_i; \theta)}{\partial\theta_j} \right|_{\theta^0} \right)^2 \right)^{\frac{1}{2}} > \epsilon\sqrt{n} \right\} \right] = 0,$$

where $I\{A\}$ is the indicator of set A.

Remark 5. Assumption 11 is the assumption required for the multivariate central limit theorem for independent non-identically distributed observations (Feller, 1971).

The following theorem states the consistency of the parametric maximum likelihood estimator.

Theorem 2.1.1. *If Assumptions 1-10 are satisfied, the maximum likelihood estimator* $\hat{\theta}_n = \left(\hat{\theta}_{1n}, \ldots, \hat{\theta}_{rn}\right)$ *is a consistent estimator of* $\theta^0 = \left(\theta_1^0, \ldots, \theta_r^0\right)$, *i.e. for all* $\zeta > 0$, $P\left(\left\|\hat{\theta}_n - \theta^0\right\| < \zeta\right) \underset{n \to +\infty}{\longrightarrow} 1$, *where* $\|.\|$ *is a norm on* Θ.

Proof. From the Taylor-Lagrange formula, from Assumption 3, for all $i = 1, \dots, n$ and for all $\theta \in \Theta$, one can write

$$\begin{aligned}\log f_i(x_i;\theta) = \log f_i(x_i;\theta^0) + \sum_{j=1}^{r} \left.\frac{\partial \log f_i(x_i;\theta)}{\partial\theta_j}\right|_{\theta^0}(\theta_j-\theta_j^0)\\ + \frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}\left.\frac{\partial^2 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\\ + \frac{1}{6}\sum_{j=1}^{r}\sum_{p=1}^{r}\sum_{q=1}^{r}\left.\frac{\partial^3 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)(\theta_q-\theta_q^0),\end{aligned}$$

where θ' belongs to the interior of the ball with center θ^0 and with radius $(\theta-\theta^0)$. By summation for i from 1 to n, inversion of sums, multiplication of both members by $1/n$ and as $(1/n)\sum_{i=1}^{n}\log f_i(x_i;\theta) = (1/n)\log L(x_1,\dots,x_n;\theta)$, that we note $(1/n)\log L(\theta)$ to lighten the notations, we have for all $\theta \in \Theta$:

$$\begin{aligned}\frac{1}{n}\log L(\theta) - \frac{1}{n}\log L(\theta^0) = \sum_{j=1}^{r}(\theta_j-\theta_j^0)\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial \log f_i(x_i;\theta)}{\partial\theta_j}\right|_{\theta^0}\\ + \frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\left[\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0} - \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]\\ + \frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]\\ + \frac{1}{6}\sum_{j=1}^{r}\sum_{p=1}^{r}\sum_{q=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)(\theta_q-\theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^3 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right]. \quad (1)\end{aligned}$$

Let us consider separately each term of the right-hand part of equation (1). From Assumption 6, we know that we have, for all $j = 1, \dots, r$,

$$\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial \log f_i(X_i;\theta)}{\partial\theta_j}\right|_{\theta^0} \xrightarrow[n\to+\infty]{P} 0.$$

From Assumption 7, we have for all $(j,p) \in \{1,\ldots,r\}^2$,

$$\frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0} - \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right) \xrightarrow[n\to+\infty]{P} 0.$$

Furthermore, for all $\theta \in \Theta$ and for all $(j,p) \in \{1,\ldots,r\}^2$,

$$\begin{aligned}\frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)& \\ = -\frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(-\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]&,\end{aligned}$$

which, from Assumption 10, is negative for all θ different from θ^0. Finally, from Assumption 8, we have for all $(j,p,q) \in \{1,\ldots,r\}^3$,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^3 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right| \xrightarrow[n\to+\infty]{P} \left|\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^3 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right)\right|$$

and this last limiting term is bounded by M for all $(j,p,q) \in \{1,\ldots,r\}^3$ thanks to Assumption 9.

Let (ζ,ε) be a vector of arbitrary positive constants. The results above allow to write three clusters of inequalities. For all ζ, for all ε, there exists n_0 such that for all n larger than n_0 and for all $(j,p,q) \in \{1,\ldots,r\}^3$, the following probabilities

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial \log f_i(X_i;\theta)}{\partial\theta_j}\right|_{\theta^0}\right| \geq \zeta^2\right),$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0} - \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right| \geq \zeta\right),$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^3 \log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right| \geq 2M\right),$$

are bounded by $\varepsilon/\left(r\left(1+r+r^2\right)\right)$. Let S denote the event involving these $r(1+r+r^2)$

inequalities:

$$\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f_i(X_i;\theta)}{\partial \theta_1} \bigg|_{\theta^0} \right| < \zeta^2, \cdots, \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^3 \log f_i(X_i;\theta)}{\partial \theta_r \partial \theta_r \partial \theta_r} \bigg|_{\theta'} \right| < 2M \right\}.$$

From the above majorations of the different probabilities, we get $P(S^*) < \varepsilon$ where S^* is the complementary of S and thus $P(S) > 1 - \varepsilon$.

Now let us study the sign of the quantity $(1/n) \log L(\theta) - (1/n) \log L(\theta^0)$ under the event S and for θ belonging to the sphere $S(\theta^0, \zeta)$ with center θ^0 and with radius ζ. Since θ belongs to $S(\theta^0, \zeta)$, there exists $j \in \{1, \ldots, r\}$ such that $|\theta_j - \theta_j^0| < \zeta$. Thus we have,

$$\left| \sum_{j=1}^{r} (\theta_j - \theta_j^0) \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f_i(x_i;\theta)}{\partial \theta_j} \bigg|_{\theta^0} \right| < \sum_{j=1}^{r} \zeta \zeta^2$$

and

$$\left| \frac{1}{2} \sum_{j=1}^{r} \sum_{p=1}^{r} (\theta_j - \theta_j^0)(\theta_p - \theta_p^0) \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f_i(x_i;\theta)}{\partial \theta_j \partial \theta_p} \bigg|_{\theta^0} \right. \right.$$
$$\left. \left. - \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E} \left(\frac{\partial^2 \log f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p} \bigg|_{\theta^0} \right) \right] \right| < \frac{1}{2} \sum_{j=1}^{r} \sum_{p=1}^{r} \zeta^2 \zeta.$$

Furthermore, as the matrix of a quadratic form is symmetric and thus diagonalizable in an orthonormal base, we have

$$\frac{1}{2} \sum_{j=1}^{r} \sum_{p=1}^{r} \left[(\theta_j - \theta_j^0)(\theta_p - \theta_p^0) \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E} \left(\frac{\partial^2 \log f_i(X_i;\theta)}{\partial \theta_j \partial \theta_p} \bigg|_{\theta^0} \right) \right] = \sum_{j=1}^{r} \gamma_j \beta_j^2,$$

where $\sum_{j=1}^{r} \beta_j^2 = \sum_{j=1}^{r} (\theta_j - \theta_j^0)^2 = \zeta^2$. From Assumption 10,

$$\sum_{j=1}^{r} \gamma_j \beta_j^2 \leq \max_j (\gamma_j) \sum_{j=1}^{r} \beta_j^2 = \max_j (\gamma_j) \zeta^2 < 0.$$

Thus,

$$\frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2\log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]\leq\max_j\,(\gamma_j)\zeta^2<0.$$

A study of the sign of the function $(1/2)r^2\zeta^3+\max_j(\gamma_j)\zeta^2$ proves that we can find ζ_0 and a positive such that for all ζ smaller than ζ_0,

$$\frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\left[\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2\log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}-\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2\log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]$$
$$+\frac{1}{2}\sum_{j=1}^{r}\sum_{p=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left.\frac{\partial^2\log f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)\right]<-a\zeta^2.$$

Lastly

$$\left|\frac{1}{6}\sum_{j=1}^{r}\sum_{p=1}^{r}\sum_{q=1}^{r}\left[(\theta_j-\theta_j^0)(\theta_p-\theta_p^0)(\theta_q-\theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\right.\right.$$
$$\left.\left.\left.\frac{\partial^3\log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right]\right|<\frac{2}{6}\sum_{j=1}^{r}\sum_{p=1}^{r}\sum_{q=1}^{r}\zeta^3M=b\;\zeta^3,$$

where $b=r^3M/3$. Gathering all the preceding inequalities, we get under S and for all θ in $S(\theta^0,\zeta)$:

$$\frac{1}{n}\log L(\theta)-\frac{1}{n}\log L(\theta^0)<r\zeta^3-a\zeta^2+b\zeta^3.$$

Assuming $\zeta<a/(r+b)$ we get under S, $(1/n)\log L(\theta)-(1/n)\log L(\theta^0)<0$ for all $\theta\in S(\theta^0,\zeta)$. Thus the event C involving for all $\theta\in S(\theta^0,\zeta)$,

$$\frac{1}{n}\log L(\theta)-\frac{1}{n}\log L(\theta^0)<0$$

is such that $P(C)\geq P(S)>1-\epsilon$.

Finally we have, for all ζ lower than $\min(\zeta_0,a/(r+b))$, the probability

$$P\left(\forall\theta\in S(\theta_0,\zeta),\frac{1}{n}\log L(\theta)-\frac{1}{n}\log L(\theta^0)<0\right)$$

tends to 1 when $n \to +\infty$. There exists $\hat{\theta}_n$ belonging to the interior of the ball $B(\theta^0, \zeta)$ i.e such that $||\hat{\theta}_n - \theta^0|| < \zeta$, such that $\log L(\theta)$ has a local maximum in $\hat{\theta}_n$. Consequently,

$$\forall \zeta \leq \min\left(\zeta_0, \frac{a}{r+b}\right),\ P\left(||\hat{\theta}_n - \theta_0|| < \zeta\right) \underset{n\to+\infty}{\longrightarrow} 1.$$

From Assumption 1 and Assumption 2, $\hat{\theta}_n$ is the maximum likelihood estimator. □

The following theorem states the asymptotic normality of the parametric maximum likelihood estimator.

Theorem 2.1.2. *If Assumptions 1-11 are satisfied, the random vector $\sqrt{n}(\hat{\theta}_n - \theta^0)$ is asymptotically normal with zero mean and covariance matrix $\left[I(\theta^0)\right]^{-1}$.*

Proof. From the Taylor-Lagrange formula and from Assumption 3, we have the following system: for all $\theta \in \Theta$, there exists $\theta' \in B^0(\theta^0, \theta - \theta^0)$, the interior of the ball with center θ^0 and radius $\theta - \theta^0$ such that for all $k = 1, \ldots, r$ we have,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f_i(x_i;\theta)}{\partial\theta_j}\bigg|_{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f_i(x_i;\theta)}{\partial\theta_j}\bigg|_{\theta^0} \\ + \sum_{p=1}^{r}(\theta_p - \theta_p^0)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\bigg|_{\theta^0} \\ + \frac{1}{2}\sum_{p=1}^{r}\sum_{q=1}^{r}(\theta_p - \theta_p^0)(\theta_q - \theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\bigg|_{\theta'}.\end{aligned}$$

But for $\theta = \hat{\theta}$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \log f_i(x_i;\theta)|_{\hat{\theta}} = 0.$$

So, for all $k = 1, \ldots, r$ we have,

$$\begin{aligned}-\sum_{p=1}^{r}(\widehat{\theta}_p - \theta_p^0)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\bigg|_{\theta^0} \\ -\frac{1}{2}\sum_{p=1}^{r}\sum_{q=1}^{r}(\widehat{\theta}_p - \theta_p^0)(\widehat{\theta}_q - \theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3 \log f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\bigg|_{\theta'} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f_i(x_i;\theta)}{\partial\theta_j}\bigg|_{\theta^0}.\end{aligned}$$

Factorizing by $\sum_{p=1}^{r}(\widehat{\theta}_p - \theta_p^0)$ and multiplying by $\sqrt{n}$,

$$-\sqrt{n}\sum_{p=1}^{r}(\widehat{\theta}_p - \theta_p^0)\left[\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \mathrm{log} f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right.$$

$$\left.+\frac{1}{2}\sum_{q=1}^{r}(\widehat{\theta}_q - \theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^3 \mathrm{log} f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left.\frac{\partial \mathrm{log} f_i(x_i;\theta)}{\partial\theta_j}\right|_{\theta^0}.$$

In terms of matrix,

$$\begin{pmatrix} a_{11n} & a_{12n} & \cdots & a_{1rn} \\ \vdots & & \ddots & \vdots \\ a_{r1n} & a_{r2n} & \cdots & a_{rrn} \end{pmatrix}\begin{pmatrix} \sqrt{n}(\widehat{\theta}_1 - \theta_1^0) \\ \vdots \\ \sqrt{n}(\widehat{\theta}_r - \theta_r^0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \left.\frac{\partial \mathrm{log} f_i(x_i;\theta)}{\partial\theta_1}\right|_{\theta^0} \\ \vdots \\ \left.\frac{\partial \mathrm{log} f_i(x_i;\theta)}{\partial\theta_r}\right|_{\theta^0} \end{pmatrix},$$

where

$$a_{jpn} = -\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \mathrm{log} f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0} - \frac{1}{2}\sum_{q=1}^{r}(\widehat{\theta}_q - \theta_q^0)\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^3 \mathrm{log} f_i(x_i;\theta)}{\partial\theta_j\partial\theta_p\partial\theta_q}\right|_{\theta'}.$$

The vectors $\nabla_\theta \mathrm{log} f_i(X_i;\theta)|_{\theta^0}$, for all $i = 1,\ldots,n$, are independent but not identically distributed with zero mean and covariance matrix $V_i(\theta^0) = \left(V_{ijp}(\theta^0)\right)_{1\le j,p\le r}$, where

$$V_{ijp}(\theta^0) = \mathrm{E}\left(\left.\frac{\partial \mathrm{log} f_i(X_i;\theta)}{\partial\theta_j}\right|_{\theta^0}\left.\frac{\partial \mathrm{log} f_i(X_i;\theta)}{\partial\theta_p}\right|_{\theta^0}\right).$$

We know from Assumption 5 that

$$V_{ijp}(\theta^0) = \mathrm{E}\left(-\left.\frac{\partial^2 \mathrm{log} f_i(X_i;\theta)}{\partial\theta_j\partial\theta_p}\right|_{\theta^0}\right)$$

and from Assumption 10 that

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}V_i(\theta^0) = I(\theta^0).$$

So, from Assumption 11 and the multivariate central limit theorem for independent non-identically distributed random variables, we get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_\theta \mathrm{log} f_i(X_i;\theta)|_{\theta^0}\xrightarrow[n\to+\infty]{d} N(0,I(\theta^0)).$$

From Assumptions 7-8, the consistency of the maximum likelihood estimator and the Slutsky's theorem, one obtains for all $(j,p)\in\{1,\dots,r\}^2$,

$$a_{jpn}\xrightarrow[n\to+\infty]{P} I_{jp}(\theta^0).$$

These convergences in probability and the weak convergence of $\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta \mathrm{log} f_i(X_i;\theta)|_{\theta^0}$ yield

$$\begin{pmatrix}\sqrt{n}(\widehat{\theta}_1-\theta_1^0)\\ \vdots \\ \sqrt{n}(\widehat{\theta}_r-\theta_r^0)\end{pmatrix}\xrightarrow[n\to+\infty]{d} N(0,\left[I(\theta^0)\right]^{-1}).$$

□

2.2 Finite number of distinct probability distribution functions

Let N be the number of distinct probability distribution functions that can be observed in the population from which the sample comes from. For $i=1,\dots,N$, let n_i be the number of observations with density $f_i(.;\theta)$ and $n=\sum_{i=1}^N n_i$ be the total number of observations. For $i=1,\dots,N$, let $\mu_i=n_i/n$ be the proportion of observations with density $f_i(.;\theta)$. One can easily prove that there exists constants $(\lambda_i)_{1\leq i\leq N}$ in $]0,1[^N$ such that for all $i=1,\dots,N$, the proportion μ_i tends to λ_i when n tends to $+\infty$. These quantities satisfy $\sum_{i=1}^N\lambda_i=1$.

Remark 6. Note that the case where there exists q such that $\lambda_q=0$ (resp. $\lambda_q=1$) corresponds to the case where there are in fact only $N-1$ distinct distributions (resp. there is only one distribution).

Let Assumptions 1-5 be the same assumptions than in the previous case, except that n is replaced by N in the Assumptions 3-5.

Assumption 12. There exists M such that for all $\theta \in \Theta$, $i = 1, \ldots, N$ and $(j, p, q) \in \{1, \ldots, r\}^3$,

$$\mathrm{E}\left(\frac{\partial^3 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p \partial\theta_q}\right)$$

exists and

$$\left|\sum_{i=1}^{N} \lambda_i \mathrm{E}\left(\frac{\partial^3 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p \partial\theta_q}\right)\right| < M.$$

Assumption 13. For all $\theta \in \Theta$, $i = 1, \ldots, N$ and $(j, p) \in \{1, \ldots, r\}^2$,

$$\mathrm{E}\left(\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right)$$

exists and the matrix $I\left(\theta^0\right) = \left(I_{jp}\left(\theta^0\right)\right)_{1 \le j,p \le r}$, where

$$I_{jp}\left(\theta^0\right) = \sum_{i=1}^{N} \lambda_i E\left(-\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right|_{\theta^0}\right),$$

is positive definite.

Remark 7. From Assumption 5, we were already sure that $I\left(\theta^0\right)$ is a positive semi-definite matrix.

The asymptotic behavior of the parametric maximum likelihood estimator in this case is given in the following two theorems.

Theorem 2.2.1. *If Assumptions 1-5 and Assumptions 12-13 are satisfied, the maximum likelihood estimator* $\hat{\theta}_n = \left(\hat{\theta}_{1n}, \ldots, \hat{\theta}_{rn}\right)$ *is a consistent estimator of* $\theta^0 = \left(\theta_1^0, \ldots,\ \theta_r^0\right)$, *i.e. for all* $\zeta > 0$, $P\left(\left\|\hat{\theta}_n - \theta^0\right\| < \zeta\right) \underset{n\to+\infty}{\longrightarrow} 1$.

Proof. Proof of Theorem 2.2.1 is similar to the proof of Theorem 2.1.1. It is sufficient to gather the observations with respect to their densities and to replace

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right|_{\theta^0}\right)$$

by

$$\sum_{i=1}^{N} \lambda_i E\left(-\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right|_{\theta^0}\right).$$

In the case of a finite number of distinct densities, the convergences in probability are given straight by the weak law of large numbers and the Slutsky's theorem. □

Theorem 2.2.2. *If Assumptions 1-5 and Assumptions 12-13 are satisfied, the random vector* $\sqrt{n}(\hat{\theta}_n - \theta^0)$ *is asymptotically normal with zero mean and covariance matrix* $\left[I(\theta^0)\right]^{-1}$.

Proof. Proof of Theorem 2.2.2 is similar to the proof of Theorem 2.1.2. It is sufficient to gather the observations with respect to their densities, replace

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right|_{\theta^0}\right)$$

by

$$\sum_{i=1}^{N} \lambda_i E\left(-\left.\frac{\partial^2 \log f_i(X_i;\theta)}{\partial\theta_j \partial\theta_p}\right|_{\theta^0}\right)$$

and use the classical multivariate central limit theorem and the Slutsky's theorem. □

References

R. A. Bradley and J. J. Gart. The asymptotic properties of ml estimators when sampling from associated populations. *Biometrika*, 49(1/2):205–214, 1962.

R. Brookmeyer, N. Blades, M. Hugh-Jones, and D. A. Henderson. The statistical analysis of truncated data: application to the sverdlovsk anthrax outbreak. *Biostatistics*, 2(2): 233–247, 2001.

K. C. Chanda. A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, 41(1/2):56–61, 1954.

W. Feller. *An Introduction to Probability Theory and Its Applications, Volume I, 3rd edition.* John Wiley & Sons, 1968.

W. Feller. *An Introduction to Probability Theory and Its Applications, Volume II, 2nd edition.* John Wiley & Sons, 1971.

S. W. Lagakos, L. M. Barraj, and V. De Gruttola. Nonparametric analysis of truncated survival data, with application to aids. *Biometrika*, 75(3):515–523, 1988.

E. L. Lehmann. *Theory of Point Estimation.* John Wiley & Sons, 1983.

A. W. Van der Vaart and J. A. Wellner. *Weak convergence and Empirical Processes.* Springer, 2000.

M. Woodroofe. Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177, 1985.